Methodology for extraction and aggregation of daily SST data and linkage to Alaska Department of Fish and Game statistical management areas.

Spatial data

Alaska’s commercial fisheries are prosecuted in federal and state waters, with many catch locations recorded by spatial strata. Two primary units for these spatial strata are the Alaska Department of Fish and Game (ADF&G) groundfish statistical areas ([www.adfg.alaska.gov/index.cfm?adfg=fishingCommercialByFishery.statmaps](http://www.adfg.alaska.gov/index.cfm?adfg=fishingCommercialByFishery.statmaps)) and the National Marine Fisheries Service (NMFS) reporting areas (alaskafisheries.noaa.gov/sites/default/files/reporting-areas.pdf). A shapefile of these areas was downloaded from the ADF&G ([soa-adfg.opendata.arcgis.com/datasets/groundfish-statistical-areas-2001](https://soa-adfg.opendata.arcgis.com/datasets/groundfish-statistical-areas-2001)) with polygons representing the state groundfish statistical areas (also called “stat6” areas). The majority of the polygons span 1° longitude x 0.5° latitude though many of the nearshore areas follow coastline contours and form less regular features. Attributes for these state polygons include additional information about the polygons related to several different fisheries and management programs including whether a polygon: occurred in state or federal waters (*STATEFED*); occurred in the Bering Sea Aleutian Islands (BSAI) or Gulf of Alaska (GOA) fishery management plan area (*FMP\_AREA\_C*); occurred within one of the sablefish management areas (*IFQ\_SABLEF*); and in which NMFS reporting area (*NMFSAREA*) the polygon occurred. An additional field (*GOA*) was added that divided the GOA region into several strata commonly used by stock assessment scientists (western GOA [WGOA], NMFSAREA 610; central GOA [CGOA], NMFSAREA 620 and 630; eastern GOA [EGOA], NMFSAREA 640 and 650).

Alaska has a clearly defined management area grid but for other areas globally with less defined structure, users could create their own spatial grid system (Bivand et al. 2013) for stratification of their fishery environment.

Temperature data

Sea surface temperature (sst) data originated from the NASA Group for High Resolution Sea Surface Temperature Level 4 analysis (JPL MUR MEaSUREs Project. 2015). These data were extracted from the NOAA CoastWatch West Coast Node ERDDAP server (<https://coastwatch.pfeg.noaa.gov/erddap/>) where they were searchable as “Multi-scale ultra-high resolution (MUR) SST Analysis fv04.1, Global, 0.01°, 2002 – present, daily.” Data from the CoastWatch site can be manually extracted using a graphical user interface that allows users to specify geographic coordinates and dates, but for querying multiple days of data, an automated data scraping script was more feasible. The data scraping script is described in conjunction with matching to the spatial management areas below. Processing of the full script (2003 – 2018) required several hundred hours of processing time.

Data processing

U.S. federal waters extend beyond the 180° meridian, resulting in both positive (west of the meridian) and negative (east of the meridian) longitudes. To simplify coding (performed via R 3.5.0 [R Core Team 2018], the temperature extraction and matching with state management areas was performed separately for the positive and negative regions (no single state statistical areas straddles the 180° meridian). Daily netcdf temperature files were downloaded and extracted (R packages **ncdf4** [Pierce 2017] and **RCurl** [Duncan Temple Lang and the CRAN team 2018]) for the North Pacific, bounded from 167°W to -130.01°W longitude and 47°N to 68° N latitude, and from 167°E to 179.9°E longitude and 47°N to 60° N. For each day, a point-in-polygon operation (R package **sp** [Pebesma and Bivand 2005; Bivand, et al., 2013]) matched all temperature data with a state statistical area. The typical 1° x 0.5° grid cell contained about 5,000 temperate records. All temperature records for each statistical area (> 4.3 million records per day) were averaged to yield a single mean and standard deviation for each statistical area and day. Before averaging across management areas, the full dataset included > 23 billion records but the final dataset was aggregated to nearly 10 million records. Matching was only performed on state statistical areas; NMFS reporting areas are larger than state statistical areas so any reporting at the NMFS reporting area level resulted from averaging the single mean value for each state statistical area contained in that NMFS area.

Occasional errors occurred during the data download process for a given day; the R script includes several quality control steps to ensure that any missing data were re-queried and rectified with the full dataset. Additional steps included extraction of Julian day, week (starting on Sunday), month, and year for each record. The **sp** package was used to extract the bounding box coordinates for each statistical area to allow the user to more easily filter areas based on the coordinates (e.g., filter out all statistical areas whose minimum latitude is less than some value).

Average depth for each statistical area was also calculated so that users could more easily subset areas that were primarily on or off the continental shelf. To maintain consistency with the temperature data, geographic coordinates for each state statistical area matched those from the temperature queries. Bathymetry data were obtained and rasterized for the same spatial ranges as for the temperature data (R package **marmap** (Pante and Simon-Bouhet 2013) and depths were extracted (R package **raster** [Hijmans 2017]) for each of the individual coordinates. Mean and standard deviations of depths were calculated for each statistical area. Areas with large standard deviations may be useful for identifying statistical areas that include the continental slope.

R scripts for data processing can be found at https://github.com/jordanwatson/ERDDAP.